Balavarca Y, Pearce KF, Norden J, Collin M, Jackson G, Holler E, Kolb HJ, Greinix H, Socie G, Toubert A, Rocha V, Gluckman E, Hromadnikova I, Sedlacek P, Wolff D, Holtick U, Dickinson AM, Bickeböller H.
Predicting survival using clinical risk score and non-HLA immunogenetics.
***Bone Marrow Transplantation* 2015, 50, 1445-1452**

Predicting survival using clinical risk scores and non-HLA immunogenetics

running title: Predicting HSCT survival

Yesilda Balavarca[1,14], Kim Pearce[2], Jean Norden[2], Mathew Collin[2], Graham Jackson[3], Ernst Holler[4], Ralf Dressel[5], Hans-Jochem Kolb[6], Hildegard Greinix[7], Gerard Socie[8], Antoine Toubert[9], Vanderson Rocha[10], Eliane Gluckman[10], Ilona Hromadnikova[11], Petr Sedlacek[12], Daniel Wolff[4], Udo Holtick[13], Anne Dickinson[2*], Heike Bickeböller[1*]

[1]Department of Genetic Epidemiology, University Medical Center, Göttingen, Germany; [2]Haematological Sciences, Institute of Cellular Medicine, Newcastle University, UK; [3] Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK; [4]Department of Hematology and Oncology, University of Regensburg, Germany; [5]Department of Cellular and Molecular Immunology, University Medical Center, Göttingen, Germany; [6]Klinikum Grosshadern, Medical Klinik III, Munich, Germany; [7]Division of Haematology, Medical University of Graz, Graz, Austria; [8]AP-HP, Saint Louis Hospital, Hematology Transplantation, Paris, France; [9]Departement d'Immunologie, Université Paris Diderot, INSERM UMRS-940, AP-HP, Paris, France; [10]EUROCORD, St Louis Hospital, Paris, France; [11]Department of Molecular Biology and Cell Pathology, Third Faculty of Medicine, Charles University Prague, Czech Republic; [12]Department of Pediatric Hematology and Oncology, Second Faculty of Medicine, Charles University Prague, Czech Republic; [13]Department I of Internal Medicine, University of Cologne, Cologne, Germany. [14]Current affiliation: Department of Preventive Oncology, National Center for Tumor Diseases and German Cancer Research Center, Heidelberg, Germany

*Anne Dickinson and Heike Bickeböller contributed equally to the work.

Corresponding author:

Name: Kim Pearce

Address: Haematological Sciences, Institute of Cellular Medicine, William Leech Building, Medical School. Newcastle University, Framlington Place, Newcastle upon Tyne, United Kingdom. NE2 4HH

Email: Kim.Pearce@ncl.ac.uk

Telephone: +44 (0) 191 208 8142; Fax: +44 (0)191 208 5524

Conflict of Interest

The authors declare no conflict of interest.

Sources of Support

This work was supported by the Marie Curie Research Training Network (MCRTN) grant CT-2004-512253: TRANSNET (European Commission), grant LSHB-CT-2007-037703: Stemdiagnostics (European Commission), the German Research Foundation (DFG), grant GRK 1034, the Marie Curie Initial Training Network (MCITN) grant 315963: CELLEUROPE, and the Deutsche José Carreras Leukämie-Stiftung.

Abstract

Previous studies of non-histocompatibility leukocyte antigen (HLA) gene single nucleotide polymorphisms (SNPs) on subgroups of patients undergoing allogeneic haematopoietic stem cell transplantation (HSCT) revealed an association with transplant outcome. This study further evaluated the association of non-HLA polymorphisms with overall survival in a cohort of 762 HSCT patients using data on 26 polymorphisms in 16 non-HLA genes. When viewed in addition to an already established clinical risk score (EBMT-score), three polymorphisms: rs8177374 in the gene for MyD88-adapter-like (MAL) (p=0.026), rs9340799 in the estrogen receptor gene (ESR) (p=0.003) and rs1800795 in interleukin 6 (IL-6) (p=0.007) were found to be associated with reduced overall survival, while the haplo-genotype (ACC/ACC) in interleukin 10 (IL-10) was protective (p=0.02). The addition of these non-HLA polymorphisms in a Cox regression model alongside the EBMT-score improved discrimination between risk groups and increased the level of prediction compared to the EBMT-score alone (gain in prediction capability for EBMT-genetic-score 10.8%). Results also demonstrated how changes in clinical practice through time have altered the effects of non-HLA analysis.

The study illustrates the significance of non-HLA genotyping prior to HSCT and the importance of further investigation into non-HLA gene polymorphisms in risk prediction.

Introduction

Haematopoietic stem cell transplantation (HSCT) is the major curative therapy for disorders of the blood and immune system. However, the rate of survival in patients with HSCT has remained at 40-60% for the last two decades, owing to post-transplant complications including infection, graft-versus-host disease (GvHD), and relapse.

Five relevant clinical factors influencing transplantation success in patients with haematological disorders, including chronic myeloid leukaemia (CML) and acute lymphocytic and acute myeloid leukaemia (ALL and AML), have been identified by the European Group for Blood and Marrow Transplantation (EBMT). These risk factors (EBMT-factors) are patient age, sibling donor/matched unrelated donor (MUD), patient-donor gender combination, stage of disease, and time from diagnosis to transplant. A clinical risk score (EMBT-score) utilising the EBMT-factors was proposed in order to aid the prediction and prevention of post-transplant complications.[1-4]

Previous genetic association studies have suggested that, besides HLA genes, non-HLA genes may play an important role in transplant outcome. To date, these studies with single nucleotide polymorphisms (SNPs) have used small subsets of patients.[5-9] Although genome-wide association studies (GWAS) have been performed,[10] no SNP genotypes have been clearly identified so far that could be used to predict outcome.

In this study, we assessed the association of candidate polymorphisms with overall survival using a large cohort of patients undergoing HSCT. The goal was to identify non-HLA SNPs with an impact on overall survival when viewed in addition to the already established EBMT-score. Prediction capability was also evaluated. Since changes in transplant protocols can affect transplant outcome, the study also took into account date of transplant.

Methods

Patients

A total of 762 patients with malignant haematological diseases, having complete data on the EBMT-factors[1,3,4] and with known date of death or last contact, were included in the study. These patients were transplanted between November 1983 and December 2005 at seven European transplant centres. The patients and donors gave informed consent to participate in the study in accordance with the Declaration of Helsinki and EBMT guidelines. The protocol was approved by the Local Research Ethics Committee at the coordinating centre (Newcastle upon Tyne, UK). Follow-up time was between 7 months and 20 years, with a median of 5-6 years. Patient and donor characteristics are presented in Table 1. Overall, death occurred in 399 patients (52%). Causes of death were relapse (41%), GvHD (18%), infection (18%), multiple organ failure (5%), acute respiratory distress syndrome (4%), veno-occlusive disease (2%), interstitial pneumonitis (1%), and others (11%). The majority of the cohort after the year 2000 had high-resolution tissue typing for HLA Class I A,B,C, and Class II DP,DQ, and DR.

Candidate non-HLA Polymorphisms

DNA was prepared from archived frozen peripheral blood mononuclear cells. Genotyping was outsourced to Kbioscience (http://www.kbioscience.co.uk) who used fluorescence-based competitive PCR technology (KASPar) and designed the assays for the SNPs based on the DNA sequence (50 bases) either side of the SNP. Genotypes were available for 743 patient-donor pairs on the following genes: CD14, CD91, C3, ESR1, GCR, HSP70-hom, IFNG, IL1RN, IL4, IL6, IL10, IL12B, IL13, LOX1, MAL, MDR1, NOD2, TNF, TNFRSF1B, and VDR (Table 2). Candidate SNPs were selected according to findings on smaller patient cohorts by our

coordinating centre in Newcastle (ESR1,[11] IFNG,[12] IL1RN,[13] IL4,[14] IL6,[15,16] IL10,[17] IL13,[18] TNF,[19] TNFRSF1B,[20] and VDR[21]). SNPs were also selected according to findings by other groups in HSCT (MAL,[22] MDR1,[23,24,25] and NOD2[26]), and according to previous disease association studies in autoimmune (CD14,[27] GCR,[28,29] HSP70-hom,[30] and IL12B[31]) or inflammatory disease (and recently found to be de-regulated in a rat model for GvHD:[32] C3,[33] LOX1,[34] and CD91[35]).

Statistical analysis

Clinical differences in patients treated up to and after year 2000 were assessed using Fisher's exact test. This division in time was chosen as transplant protocols were changed at that time with the introduction of Imatinib and a subsequent increase in survival rate.[2]

Biallelic SNPs were considered under the additive, dominant, and recessive modes of inheritance[10] (Supplementary Section A). In the models, each SNP was used with the mode demonstrating the strongest association with survival.

The genes HSP70-hom, IL12B, and MDR1 and GCR-haplotype were excluded from the analysis due to missing genotypes (missingness>33%). A total of 26 polymorphisms from 16 genes were available for analysis (Table 2).

A power of 80% in the cohort with 52% deaths was achieved for SNPs in an additive mode for sample sizes n≥300, minor allele frequency (MAF)>15%, and an expected hazard ratio HR≥1.50. In a dominant mode, this power was achieved for n≥400. In a recessive mode, this was attained for n≥500, MAF≥25%, and HR≥2.00. A few SNPs with a MAF<10% (i.e. IL6 rs1800796, LOX1 rs11053646, and the three SNPs in NOD2) had a very low power at all settings; even so, we retained these SNPs for further analysis.

Association of risk score for overall survival

The EBMT risk score (EBMT-score) was derived by a summation procedure of the EBMT-factors.[1,3,4] The EBMT-score was subsequently implemented in analyses on an ordinal scale [low to high risk 0-7]. The additional effect of each individual polymorphism was evaluated using separate Cox regression models. The likelihood ratio test (LRT) was applied to compare a Cox regression model including the EBMT-score and one polymorphism to a model including only the EBMT-score; a nominal p-value of 0.05 was used. Analyses were performed for the cohort as a whole as well as for several subgroups.

Using stepwise selection, we built an additional Cox regression model to establish whether multiple non-HLA polymorphisms improved the model when added in with the EBMT-score (α=0.05 for variable entry and α=0.10 for variable removal[36]). All available polymorphisms were chosen as candidates in this procedure. Hazard ratios (HRs) and 95% confidence intervals (95% CI) were reported.

A new risk score (EBMT-genetic-score) was developed using this last Cox regression model. The score was derived by summing up the clinical EBMT-score and the genetic score points given for the polymorphisms. The latter were obtained by dividing the respective regression coefficients by the coefficient of the EBMT-score and rounding to the nearest integer[37] (Supplementary Section B).

Assessment of prediction

Two statistical approaches were used to assess the risk score prediction capability: the concordance index (C-index)[3,4,38-40] and the R-square measure of the gain in prediction (R^2).[41] The C-index measures the agreement between risk score and observed survival time. A higher risk score should correspond to a shorter observed survival time. A C-index=0.5

implies that a risk score has no predictive discrimination, whereas C-index=1 implies maximum predictive discrimination. The U-statistic[40] was also used to test whether the EBMT-genetic-score was better than the EBMT-score as regards agreement with observed survival time. In addition, R^2 was used to quantify the improvement in prediction[41] of the new EBMT-genetic-score over the EBMT-score. This measure is a combination of the 0.632 bootstrap estimate of prediction error[42] and the explained variation using Schoenfeld residuals[43-44] (see Supplementary Section C). Larger values of R^2 (>0%) mean that the EBMT-genetic-score correctly predicts outcome more often than the EBMT-score. R^2=0% means that both scores have equivalent predictive ability and R^2=100% means that the EBMT-genetic-score has perfect predictive ability (i.e. the predicted and the actual outcomes always agree).

Results

Clinical characteristics

Sixty percent of the patients underwent HSCT after the year 2000. There was evidence that, after 2000, transplants involved more patients and donors over 40 years of age, more HLA-matched unrelated donors, more lymphoma, more donor cells from peripheral blood, more later stage disease, more T-cell depletion, and more reduced-intensity conditioning (RIC) (Table 1).

There was no significant centre effect on overall survival (likelihood ratio test, P value= 0.49).

Association of the EBMT risk score and single polymorphisms with overall survival

The EBMT-score[1,3,4] was significantly associated with overall survival (HR=1.16, 95% CI=1.09-1.24, p<0.001).

The top ten candidate polymorphisms associated with overall survival in the whole cohort, while controlling for EBMT-score, are given in Table 3. The IL10 haplotype in donors demonstrated the lowest p-value and the presence of ACC/ACC was protective (HR=0.48, 95% CI=0.29-0.80, LRT p-value=0.002). This haplotype was also the only polymorphism significantly associated with survival in all subgroups (Supplementary Table 1). Nine of the ten polymorphisms were also highest ranked amongst patients without T-cell depletion and with myeloablative conditioning regimens. IL10 haplotype, IL10 rs1800896(G), IL4 rs2243250(T), and IL6 rs1800797(A) in donors, ESR1 rs2234693(C) and GCR rs33388(T) in patients were observed in over 50% of the assessed clinical subgroups.

Association and prediction of multiple polymorphisms with overall survival

After stepwise selection, the final model contained the EBMT-score and four selected polymorphisms (n=419 due to missing genotypes). The presence of haplo-genotype ACC/ACC of IL10 in donors was protective against patient death (HR=0.49, 95% CI=0.26-0.89, p-value=0.020). The risk of death increased with an increased number of T alleles in MAL rs8177374 in patients (additive, HR=1.34, 95% CI=1.04-1.74, p-value=0.026), with the presence of allele G in ESR1 rs9340799 in patients (dominant, HR=1.52, 95% CI=1.15-2.01, p-value=0.003), and with the increased number of C alleles in IL6 rs1800795 in donors (additive, HR=1.29, 95% CI=1.07-1.55, p-value=0.007) (Table 4, Supplementary Figure 1).

Three out of four of these polymorphisms were significantly associated with death due to relapse. In addition, MAL rs8177374 was associated with death due to GvHD and IL6 rs1800795 was associated with death due to infection (Supplementary Table 2).

When the multivariate Cox regression model (Table 4) was compared to a model containing the EBMT-score alone, the estimated R^2 for gain in prediction indicated a 5.1% gain in prediction ability by adding the four polymorphisms (separately) to the EBMT-score.

Comparing EBMT-genetic-score and EBMT-score

The new EBMT-genetic-score, derived through summing the individual risk score values (Supplementary Section B), ranged from 1 to 15. The scores were grouped into five distinct categories according to the observed groupings of the Kaplan-Meier survival curves (results not shown). The Cox regression model, with five ordered categories of the EBMT-genetic-score, revealed increasing hazard ratios with increasing EBMT-genetic-score (n=419, Table 5). The risk values of the EBMT-genetic-score displayed a clearer separation of the survival curves when compared to values of the EBMT-score (Supplementary Figure 2).

Kaplan-Meier curves for the EBMT-genetic-score were plotted to establish if the EBMT-genetic score is appropriate for the clinical subgroups: disease diagnosis, sibling donor/matched unrelated donor, T-cell depletion, conditioning regimen, and year of transplantation. The plots consistently demonstrated that higher risk scores corresponded to lower survival (Supplementary Figure 3).

For the whole cohort, the higher EBMT-genetic-score corresponded to shorter observed survival times compared to the EBMT-score (U-statistic p-value<0.001, Table 6). This also proved to be the case in subgroups of patients transplanted before and after 2000. Estimation of the gain in prediction ability indicated that there was benefit in utilising the single EBMT-genetic-score (R^2=10.8%, Table 6) over the previous model containing the EBMT-score and four separate polymorphisms (R^2=5.1%).

Kaplan Meier survival curves for EBMT-score and EBMT-genetic-score before and after 2000 (n=419) appear in Figures 1a-1d. Compared to the EBMT-score, the EBMT-genetic-score better discriminates the survival curves and a higher score consistently corresponds to a lower survival probability. It was also apparent that, when using the EBMT-genetic-score, those treated after 2000 had improved chances of survival in comparison with those treated before 2000 (Figures 1c, 1d). For those with the lowest scores (1-6), 3-year survival was 85% and 95% for patients treated before and after 2000, respectively; for those with the highest scores (13-15), 1-year survival was 15% and 42% for patients treated before and after 2000, respectively.

Discussion

The aim of this work was to study the effect of non-HLA polymorphisms on overall survival of HSCT patients in addition to the EBMT-score. In our study, the IL10 promoter haplo-genotype ACC/ACC in donors was one of the most important polymorphisms associated with improved overall survival. IL10 is an important cytokine in the regulation of the immune response. However, it can have a stimulatory effect on B cells, increasing MHC class II expression and antibody production. IL10 haplotypes have been shown to correlate with IL10 protein production,[45] with the GCC haplotype being associated with the highest IL10 production.

The ATA/ACC genotype has been identified as a protective factor for overall survival in CML patients with sibling donors.[9] SNP rs1800872(A) and haplotype ACC in patients have been shown to demonstrate a strong association with severe acute GvHD (aGvHD III-IV) in patients with matched related donors.[7,10] In our whole cohort, the IL10 promoter haplo-genotype ACC/ACC in donors proved to be significantly associated with increased overall survival,

whereas SNP rs1800872(A) in patients revealed only borderline association. The ACC/ACC genotype in the donor is associated with intermediate production of IL10 and the ACC haplotype has been shown to be protective in aspergillosis.[46] The AA genotype of IL10 rs1800872 in the patient is associated with a decreased risk of aGvHD[47] and an increased risk of non-relapse mortality; this was confirmed in follow-up GWAS studies.[10,48] In addition, the presence of the GG genotype of IL10 rs1800896 in the patient was found to be associated with the risk of chronic GvHD (cGvHD).[49]

Furthermore, we discovered that MAL rs8177374(T) in patients is associated with reduced overall survival whereas parallel smaller studies have reported that the presence of the T allele in donors resulted in less cGvHD and a reduction in transplant-related mortality. In addition, the presence of the T allele in patients was associated with an increased risk of relapse.[50] Our study revealed a strong association between the presence of the T allele and relapse and an increased risk of death in patients who relapsed. Moreover, the T allele was strongly associated with GvHD. The T allele is regarded in the literature as the inflammatory allele, and T-heterozygous individuals have increased protection from infection. Interestingly, another study revealed that patients transplanted from donors with the T allele have a lower incidence of fungal infections, aGvHD, and improved overall survival.[22] MAL protein was originally identified in intermediate and late stages of T-lymphocyte differentiation[51] and the MAL mRNA expression was also found to be related with differentiation in urothelial cells, neuronal cells,[52,53] and oesophageal epithelium.[54] The MAL-A variant containing all four exons is abundantly expressed in peripheral blood lymphocytes[55] and positively expressed in the gastrointestinal tract, respiratory tract, and haematopoietic system. MAL is important in the innate immune response; it is involved in

279 Toll-like receptor signalling,[56] so could be important in the development of aGVHD in the
280 recipient.
281 Our study also indicated that ESR1 rs9340799(G) in patients is associated with reduced
282 overall survival. The G allele has been previously reported as associated with reduced overall
283 survival and increased risk of aGvHD in patients with HLA-matched siblings.[11] ESR1 is thought
284 to inhibit IL6 production.[57]
285 Furthermore, we found IL6 rs1800795(C) in donors to be associated with reduced overall
286 survival. The G allele in patients and/or donors has been reported by GWAS studies as a risk
287 factor for both aGvHD and cGvHD in patients with HLA-matched related[10,21] and unrelated[10]
288 donors. In our study, there was evidence of an increased risk of aGvHD in patients
289 transplanted from donors with the C allele present. Among these patients, those
290 transplanted before 2000 had evidence of reduced survival compared to those transplanted
291 later (Supplementary Figure 4a), possibly as a result of a higher rate of standard
292 myeloablative conditioning before 2000.
293 The G allele of this SNP is reported to correlate with higher serum IL6 levels in systemic-
294 onset juvenile chronic arthritis.[58] However, the CC genotype has been associated with higher
295 levels of IL6 in polymyalgia rheumatica.[59] It could therefore be that increased levels of IL6 in
296 our cohort exacerbated the inflammatory milieu, leading to increased transplant-related
297 complications such as infection and poorer survival.
298 We have also demonstrated that polymorphisms modelled via Cox regression, either in a
299 joint model with the EBMT-score or combined with the latter as a single EBMT-genetic-score
300 factor, contribute to a better discrimination of the risk groups (C-index) and increase the
301 prediction of survival (R^2) compared to the EBMT-score alone. Changes in HSCT clinical
302 protocols during 2000 greatly improved patient survival. The consideration of an EBMT-

genetic-score highlighted an improvement in survival, especially for those at higher risk of death. More transplants from MUDs were performed after 2000 and it could be argued that the improvement in survival of patients with high-risk non-HLA genotypes is due to the improved quality of HLA matching after 2000. However, no difference in survival was evident between siblings or MUD patients transplanted after 2000 (Supplementary Figure 4b).

In addition, a recent review has revealed that other pre-transplant clinical factors (e.g., CMV status and Karnofsky performance score) also play a role in survival and could be used alongside EBMT-score[4].

In conclusion, we hypothesise that implementing risk scores for pre-transplant risk assessment from clinical and genetic factors enhances the prediction of overall survival for patients undergoing HSCT. The potential of considering non-HLA polymorphisms in pre-transplant risk assessment is evident with the promising results for polymorphisms in genes IL10, MAL, ESR1 and IL6. Further investigations into pre-transplant risk assessment could also include other potential predictors such as mRNA and microRNA expression.[60]

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

This work was supported by the Marie Curie Research Training Network (MCRTN) grant CT-2004-512253: TRANSNET (European Commission), grant LSHB-CT-2007-037703: Stemdiagnostics (European Commission), the German Research Foundation (DFG), grant GRK

1034, the Marie Curie Initial Training Network (MCITN) grant 315963: CELLEUROPE, and the Deutsche José Carreras Leukämie-Stiftung. We would like to thank Andrew Entwistle for his support in reviewing and proofreading the manuscript.

Supplementary information is available at Bone Marrow Transplantation's website.

References

1. Gratwohl A, Hermans J, Goldman JM, Arcese W, Carreras E, Devergie A *et al*. Risk assessment for patients with chronic myeloid leukaemia before allogeneic blood or marrow transplantation. *The Lancet* 1998; **352**: 1087-1092.

2. Gratwohl A, Brand R, Apperley J, Crawley C, Ruutu T, Corradini P *et al*. Allogeneic hematopoietic stem cell transplantation for chronic myeloid leukemia in Europe 2006: transplant activity, long-term data and current results. An analysis by the Chronic Leukemia Working Party of the European Group for Blood and Marrow Transplantation (EBMT). *Haematologica* 2006; **91**: 513-521.

3. Gratwohl A, Stern M, Brand R, Apperley J, Baldomero H, de Witte T *et al*. Risk score for outcome after allogeneic hematopoietic stem cell transplantation: a retrospective analysis. *Cancer* 2009; **115**: 4715-4726.

4. Gratwohl A. The EBMT Risk Score. *Bone Marrow Transpl* 2012; **47**: 749-756.

5. Mullally A, Ritz J. Beyond HLA: the significance of genomic variation for allogeneic hematopoietic stem cell transplantation. *Blood* 2007; **109**: 1355-1362.

6. Ambruzova Z, Mrazek F, Raida L, Jindra P, Vidan-Jeras B, Faber E *et al*. Association of IL6 and CCL2 gene polymorphisms with the outcome of allogeneic haematopoietic stem cell transplantation. *Bone Marrow Transpl* 2009; **44**: 227-235.

7. Dickinson AM, Middleton PG, Rocha V, Gluckman E, Holler E. Genetic polymorphisms predicting the outcome of bone marrow transplants. *Brit J Haematol* 2004; **127**: 479-490.

8. Dickinson A. Non-HLA genetics and predicting outcome in HSCT. *Int J Immunogenet* 2008; **35**: 375-380.

9. Dickinson AM, Pearce KF, Norden J, O'Brien SG, Holler E, Bickeböller H *et al*. Impact of genomic risk factors on outcome after hematopoietic stem cell transplantation for patients with chronic myeloid leukemia. *Haematologica* 2010; **95**: 922-927.

10. Chien JW, Zhang XC, Fan W, Wang H, Zhao LP, Martin PJ *et al*. Evaluation of published single nucleotide polymorphisms associated with acute GVHD. *Blood* 2012; **119**: 5311-5319.

11. Middleton PG, Norden J, Cullup H, Cavet J, Jackson GH, Taylor PR *et al*. Oestrogen receptor alpha gene polymorphism associates with occurrence of graft-versus-host disease and reduced survival in HLA-matched sib-allo BMT. *Bone Marrow Transpl* 2003; **32**: 41-47.

12. Pravica V, Asderakis A, Perrey C, Hajeer A, Sinnott PJ, Hutchinson IV. In Vitro Production of IFN-Gamma Correlates with CA Repeat Polymorphism in the Human IFN-Gamma Gene. *Eur J Immunogenet* 1999; **26**: 1-3.

13. Hurme M, Santtila S. IL-1 receptor antagonist (IL-1Ra) plasma levels are co-ordinately regulated by both IL-1Ra and IL-1beta genes. *Eur J Immunol* 1998; **28**: 2598-2602.

14. Walley AJ, Cookson WO. Investigation of an Interleukin-4 promoter polymorphism for associations with asthma and atopy. *J Med Genet* 1996; **33**: 689-692.

15. Cavet J, Dickinson AM, Norden J, Taylor PR, Jackson GH, Middleton PG. Interferon-gamma and interleukin-6 gene polymorphisms associate with graft-versus-host disease in HLA-matched sibling bone marrow transplantation. *Blood* 2001; **98**: 1594-1600.

16. Leland JY, KyungAh I, Borg B, Yang H, Liang TJ. Interleukin-6 (IL-6) haplotypes and the response to therapy of chronic hepatitis C virus infection. *Genes Immun* 2009; **10**: 365-372.

17. Morse HR, Olomolaiye OO, Wood NA, Keen LJ, Bidwell JL. Induced heteroduplex genotyping of TNFα, IL-1β, IL-6 and IL-10 polymorphisms associated with transcriptional regulation. *Cytokine* 1999; **11**: 789-795.

18. Graves PE, Kabesch M, Halonen M, Holberg C J, Baldini M, Fritzsch C *et al*. A cluster of seven tightly linked polymorphisms in the IL-13 gene is associated with total serum IgE levels in three populations of white children. *J Allergy Clin Immunol* 2000; **105**: 506-513.

19. Imboden M, Nieters A, Bircher AJ, Brutsche M, Becker N, Wjst M *et al.* Cytokine gene polymorphisms and atopic disease in two European cohorts. (ECRHS-Basel and SAPALDIA). *Clin Mol Allergy* 2006; **4**: 9.

20. Stark GL, Dickinson AM, Jackson GH, Taylor PR, Proctor SJ, Middleton PG. Tumour necrosis factor receptor type II 196 M/R genotype correlates with circulating soluble receptor levels in normal subjects and graft-versus-host disease after sibling allogeneic bone marrow transplantation. *Transplantation* 2003; **76**: 1742-1749.

21. Middleton PG, Cullup H, Dickinson AM, Norden J, Jackson GH, Taylor PRA *et al*. Vitamin D receptor gene polymorphism associates with graft-versus-host disease and survival in HLA-Matched sibling allogeneic bone marrow transplantation. *Bone Marrow Transpl* 2002; **30**: 223-228.

22. Rocha V, Porcher R, Kabbara, N, Peffault de Latour R *et al*. Mutant TIRAP gene polymorphism is associated with outcomes in HLA genoidentical bone marrow

transplants recipients with leukemia. *Blood (ASH Annual Meeting Abstracts)* 2007; Abstract 327.

23. Hoffmeyer S, Burk O, von Richter O, Arnold HP, Brockmoeller J, Johne A *et al*. Functional polymorphisms of the human multidrug-resistance gene: multiple sequence variations and correlation of one allele with P-glycoprotein expression and activity in vivo. *Proc Natl Acad Sci USA* 2000; **97**: 3473-3478.

24. Kim RB, Leake BF, Choo EF, Dresser GK, Kubba SV, Schwarz UI *et al*. Identification of functionally variant MDR1 alleles among European Americans and African Americans. *Clin Pharmacol Ther* 2001; **70**: 189-199.

25. Kim DH, Park JY, Sohn SK, Lee NY, Suh JS, Lee KB. The association between multidrug resistance-1 gene polymorphisms and outcomes of allogeneic HLA-identical stem cell transplantation. *Haematologica* 2006; **91**: 848-851.

26. Holler E, Rogler G, Herfarth H, Brenmoehl J, Wild PJ, Hahn J *et al*. Both donor and recipient NOD2/CARD15 mutations associate with transplant-related mortality and GvHD following allogeneic stem cell transplantation. *Blood* 2004; **104**: 889-894.

27. Klein W, Tromm A, Griga T, Fricke H, Folwaczny C, Hocke M *et al*. A polymorphism in the CD14 gene is associated with Crohn disease. *Scand J Gastroentero* 2002; **37**: 189-191.

28. Stevens A, Ray DW, Zeggini E, John S, Richards HL, Griffiths CE *et al*. Glucocorticoid sensitivity is determined by a specific glucocorticoid receptor haplotype. *Clin Endocrinol Metab* 2004; **89**: 892-897.

29. Derijk RH, Schaaf MJ, Turner G, Datson NA, Vreugdenhil E, Cidlowski J *et al*. A human glucocorticoid receptor gene variant that increases the stability of the glucocorticoid

receptor beta-isoform mRNA is associated with rheumatoid arthritis. *Rheumatology* 2001; **28**: 2383-2388.

30. Bogunia-Kubik K, Lange A. HSP70-hom gene polymorphism in allogeneic hematopoietic stem-cell transplant rescipients correlates with the development of acute graft-versus-host disease. *Transplantation* 2005; **79**: 815-820.

31. Cargill M, Schrodi SJ, Chang M, Garcia VE, Brandon R, Callis KP *et al*. A large-scale genetic association study confirms IL12B and leads to the identification of IL23R as psoriasis-risk genes. *Am J Human Genet* 2007; **80**: 273-290.

32. Novota P, Zinocker S, Norden J, Wang WN, Sviland L, Opitz L *et al*. Expression profiling of major histocompatibility and natural killer complex genes reveals candidates for controlling risk of graft versus host. *PLoS One* 2011; **6**: e16582.

33. Park KH, Fridaley BL, Ryu E, Tosakulwong N, Edwards AO. Complement component 3 (C3) haplotypes and risk of age-related macular degeneration. *Invest Ophth Vis Sci* 2009; **50**: 3386-3393.

34. Wang L, Yanuck D, Beecham A, Gardener H, Slifer S, Blanton SH *et al*. A candidate gene study revealed sex-specific association between the OLR1 gene and carotid plaque. *Stroke* 2011; **42**: 588-592.

35. Chalmers KA, Barker R, Passmore PA, Panza F, Seripa D, Solfrizzi V *et al*. LRP-1 variation is not associated with risk of Alzheimer's disease. *Int J Mol Epidemiol Genet* 2010; **1**: 104-113.

36. Labopin M, Iacobelli S. *Statistical guidelines for EBMT*. The European Group for Blood and Marrow Transplantation. Available at:

http://www.ebmt.org/1WhatisEBMT/Op_Manual/OPMAN_StatGuidelines_oct2003.pdf. Accessed August 12, 2011.

37. Parmar M, Machin D. *Survival Analysis. A practical approach*. Chichester, England and New York: J Wiley & Sons, 1995.

38. Harrell FE, Jr., Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. *J Am Med Assoc* 1982; **247**: 2543-2546.

39. Harrell FE, Jr., Lee KL, Califf RM, Pryor DB, Rosati RA. Regression modelling strategies for improved prognostic prediction. *Statistics in medicine* 1984; **3**: 143-152.

40. Harrell F and with contributions from many other users. *Hmisc: Harrell miscellaneous*; rcorrp.cens, method=2. R package version 3.5-2; 2008.

41. Balavarca Y. *Assessing prediction error of genetic variants in Cox regression models.* Doctoral thesis, available at: http://webdoc.sub.gwdg.de/diss/2012/balavarca/. Accessed April 30, 2012.

42. Efron B. Estimating the error rate of a prediction rule: improvement on cross-validation. *J Am Stat Assoc* 1983; **78**: 316-331.

43. Schoenfeld D. Partial residuals for the proportional hazards regression model. *Biometrika* 1982; **69**: 239-241.

44. O'Quigley J, Xu R. *Explained variation in Cox regression*. Handbook of Statistics in Clinical Oncology. NewYork: Marcel. Dekker, Inc.: 397-410, 2001.

45. Turner DM, Williams DM, Sankaran D, Lazarus M, Sinnot PJ, Hutchinson IV. An investigation of polymorphism in the interleukin-10 gene promoter. *Eur J Immunogenet* 1997; **24**: 1-8

46. Seo KW, Kim DH, Sohn SK, Lee NY, Chang HH, Kim SW *et al*. Protective role of interleukin-10 promoter gene polymorphisms in the pathogenesis of invasive pulmonary aspergillosis after allogeneic stem cell transplantation. *Bone Marrow Transplant* 2005; **36**: 1089-1095.

47. Lin MT, Storer B, Martin PJ, Tseng LH, Gooley T, Chen PJ *et al.* Relation of an interleukin-10 promoter polymorphism to graft-versus-host disease and survival after hematopoietic-cell transplantation. *New Engl J Med* 2003; **349**: 2201-2210.

48. Hansen JA, Chien JW, Warren EH, Zhao LP, Martin PJ. Defining genetic risk for GVHD and mortality following allogeneic hematopoetic stem cell transplantation *Curr Opin Hematol* 2010; **17**: 483-492.

49. Rocha V, Franco RF, Porcher R, Bittencourt H, Silva WA, Jr., Latouche A *et al.* Host defense and inflammatory gene polymorphisms are associated with outcomes after HLA-identical sibling bone marrow transplantation. *Blood* 2002; **100**: 3908-3918.

50. Holtick U, Norden J, Jackson GJ, Holler E, Hromadnikova I, Sedlacek P *et al*. The TLR adapter Mal/TIRAP protein variant leu 180 influences graft-versus-leukaemia and graft-versus-host reactions after allogeneic haematopoietic stem cell transplantation. *34th Annual Meeting of the European Group for Blood and Marrow Transplantation* 2008; **41**(Suppl.1): S217.

51. Alonso MA, Weissman SM. cDNA cloning of MAL, a hydrophobic protein associated with human T-cell differentiation. *Proc Natl Acad Sci USA* 1987; **84**: 1997-2001.

52. Liebert M, Hubbel A, Chung M, Wedemeyer G, Lomax MI, Hegeman A *et al*. Expression of *mal* is associated with urothelial differentiation in vitro: identification

by differential display reverse-transcriptase polymerase chain reaction. *Differentiation* 1997; **61**: 177-185.

53. Wakeman JA, Heath PR, Pearson C, Andrews PW. MAL mRNA is induced during the differentiation on human embryonal carcinoma cells into neurons and is also localised within specific regions of the human brain. *Differentiation* 1997; **62**: 97-105.

54. Mimori K, Nishida K, Nakamura Y, Ieta K, Yoshikawa Y, Sasaki A *et al.* Loss of MAL expression in precancerous lesions of the esophagus. *Ann Surg Oncol* 2007; **14**: 1670-1677.

55. Rancano C, Rubio T, Correas I, Alonso M. Genomic structure and subcellular localisation of MAL, a human T-cell-specific proteolipid protein. *J Biol Chem* 1994; **269**: 8159-8164.

56. Akira S, Takeda K. Toll-like receptor signalling. *Nat Rev Immunol* 2004; **4**: 499-511.

57. Girasole G, Jilka RL, Passeri G, Boswell S, Boder G, Williams DC *et al*. 17 beta-estradiol inhibits interleukin-6 production by bone marrow-derived stromal cells and osteoblasts in vitro: a potential mechanism for the antiosteoporotic effect of estrogens. *J Clin Invest* 1992; **89**: 883-891.

58. Fishman D, Faulds G, Jeffery R, Mohamed-Ali V, Yudkin JS, Humphries S *et al.* The effect of novel polymorphisms in the interleukin-6 (IL-6) gene on IL-6 transcription and plasma IL-6 levels, and an association with systemic-onset juvenile chronic arthritis. *J Clin invest* 1998; **102**: 1369-1376.

59. Boiardi L, Casali B, Farnetti E, Pipitone N, Nicoli D, Cantini F *et al.* Relationship between interleukin 6 promoter polymorphism at position -174, IL-6 serum levels,

512 and the risk of relapse/recurrence in polymyalgia rheumatica. *J Rheumatol* 2006; **33**:
513 703-708.

514 60. Xiao B, Wang Y, Li W, Baker M, Guo J, Corbet K *et al*. Plasma microRNA signature as a
515 noninvasive biomarker for acute graft-versus-host disease. *Blood* 2013; **122**: 3365-
516 3375.

517

Figure 1 Kaplan-Meier curves of patients undergoing HSCT up to and after the year 2000 (n=419) for 5 years follow-up a) EBMT-scores before 2000, b) EBMT-scores after 2000, c) EBMT-genetic-scores before 2000 and d) EBMT-genetic-scores after 2000.

Supplementary Figure 1 Kaplan-Meier survival curves (for 5years follow-up) of patients undergoing HSCT for each polymorphism appearing in the Cox regression model (n=419) a) IL10 haplotype – donor b) MAL rs8177374– patient c) ESR1 rs9340799 - patient d) IL6 rs1800795-donor.

Supplementary Figure 2 Kaplan-Meier survival curves (for 5 years follow-up) of patients undergoing HSCT a) EBMT-score (n=743), b) EBMT-genetic-score (n=419).

Supplementary Figure 3 Kaplan-Meier survival curves (for 5 years follow-up) of patients undergoing HSCT according to the EBMT-genetic-score, by various clinical subgroups.

Supplementary Figure 4 Kaplan-Meier survival curves (for 5 years follow-up) of patients undergoing HSCT according to year of transplant, by type of donor

Supplementary Table 1 Top ten polymorphisms from the whole cohort that remained as top polymorphisms in the subgroup analyses for association with overall survival.

Supplementary Table 2 P-values for the association of selected polymorphisms with different causes of death (n=419), using a Cox regression model including a single polymorphism and the EBMT-score.

Supplementary Sections A, B, C – Additional information describing: a) the method of coding the biallelic SNPs (additive, dominant and recessive), b) an example of risk score calculation c) derivation of the R-square measure of the gain in prediction.

Table 1. Clinical characteristics of patients and donors prior to HSCT (n=762)

Factors	Categories	No.	(%)	Year of Transplant ≤ 2000		>2000	
				No.	(%)	No.	(%)
No. of patient/donor		762		302		460	
Age of patient at transplantation (years)[a,b]	<20 years	32	(4)	22	(7)	10	(2)
	20-40 years	346	(46)	171	(57)	175	(38)
	>40 years	384	(50)	109	(36)	275	(60)
Age of donor (years)[b,g]	<20 years	34	(4)	25	(8)	9	(2)
	20-40 years	424	(56)	169	(56)	255	(55)
	>40 years	286	(38)	98	(33)	188	(41)
	NA[c]	18	(2)	10	(3)	8	(2)
Patient/donor gender combination[a]	Male/female	164	(22)	64	(21)	100	(22)
	Other	598	(78)	238	(79)	360	(78)
Type of donor[a,b]	Sibling	395	(52)	188	(62)	207	(45)
	HLA-matched unrelated(MUD)	367	(48)	114	(38)	253	(55)
Haematological disease[b,g]	Acute leukaemia (AL)	391	(51)	154	(51)	237	(52)
	Chronic myeloid leukaemia (CML)	172	(23)	108	(36)	64	(14)
	Lymphoma	112	(15)	28	(9)	84	(18)
	other diagnoses[d]	87	(11)	12	(4)	75	(16)
Source of stem cells[b]	Bone Marrow	356	(47)	247	(82)	109	(24)
	Peripheral blood	388	(51)	48	(16)	340	(74)
	Both sources	3	(0)	2	(<1)	1	(<1)
	NA[c]	15	(2)	5	(2)	10	(2)
Patient/donor CMV status[g]	Negative/Negative	227	(30)	88	(29)	139	(30)
	Other	511	(67)	201	(67)	310	(67)
	NA[c]	24	(3)	13	(4)	11	(3)
Stage of disease at transplantation[a,b,g]	Early	328	(43)	166	(55)	162	(35)
	Intermediate	196	(26)	63	(21)	133	(29)
	Late	238	(31)	73	(24)	165	(36)
Time from diagnosis to transplant[a]	≤12 months	439	(58)	175	(58)	264	(57)
	>12 months	323	(42)	127	(42)	196	(43)
T-cell depletion[b]	T-cell depletion	263	(35)	68	(23)	195	(42)
	No T-cell depletion	499	(65)	234	(77)	265	(58)
Conditioning regimen[b]	Standard myeloablative	503	(66)	263	(87)	240	(52)
	Reduced-Intensity (RIC)	259	(34)	39	(13)	220	(48)
Transplantation center[b]	Vienna/Prague[e]	225	(30)	73	(24)	152	(33)
	Regensburg/Munich[f]	217	(28)	72	(24)	145	(31)
	Newcastle	204	(27)	123	(41)	81	(18)
	Rostock	73	(9)	23	(8)	50	(11)
	Paris	43	(6)	11	(3)	32	(7)

[a] Clinical EBMT factor[3].

[b] P value ≤ 0.05, Fisher-exact-test for differences between patients treated up to and after 2000. [c] NA=not available.

[d] Other diagnoses included: plasma cell neoplasia, myelodysplasia syndrome, and chronic myelomonocytic leukaemia.

[e,f] The centers of Vienna (29% patients) and Prague (1% patients) worked in close collaboration with each other. Hence, both centers used similar treatment procedures. Likewise, there were collaborations between the centers of Regensburg (27.5% patients) and Munich (0.5% patients).

[g] Factor showing significant association with overall survival (log-rank test, P value ≤ 0.05).

Table 2. Descriptives of non-HLA SNPs, and other polymorphisms[g]. (n=743)

Gene name		Chr.[a]	SNP rs number	Allele pair	MA[b]	MAF % [c]		Sample Size [d]	
						patient	donor	patient	donor
Cluster determinant 14	(CD14)	5	rs2569190	G/A	A	49	47	536	542
Cluster determinant 91	(CD91)	12	rs1799986	C/T	T	15	16	514	560
Complement component 3	(C3)	19	rs2230199	C/G	G	22	22	523	550
Estrogen receptor 1	(ESR1)	6	rs2234693	C/T	C	46	44	646	635
			rs9340799	G/A	G	40	37	646	644
Glucocorticoid receptor	(GCR)	5	rs33389	C/T	T	14	15	521	561
			rs33388	T/A	T	44	46	527	562
			rs6198	G/A	G	17	17	533	561
Heat shock protein 70 Hom	(HSP70-hom) [e]	6	rs2075800	G/A	A	35	32	433	417
			rs2227956	T/C	C	18	19	495	488
IL1 receptor antagonist	(IL1RN)	2	rs419598	T/C	C	24	24	663	671
Interleukin 4 Hom	(IL4)	5	rs2243250	T/C	T	15	16	617	618
Interleukin 6	(IL6)	7	rs1800797	G/A	A	39	39	561	593
			rs1800796	C/G	C	7	6	582	618
			rs1800795	G/C	C	42	39	664	678
Interleukin 10	(IL10)	1	rs1800896	G/A	G	47	46	609	619
			rs1800872	A/C	A	24	26	620	626
Interleukin 12B	(IL12B) [e]	5	rs3212227	A/C	C	21	20	440	471
Interleukin 13	(IL13)	5	rs1800925	C/T	T	19	18	543	592
			rs20541	A/G	A	21	22	577	596
			rs1881457	C/A	C	19	19	584	613
Oxidized Low-Density Lipoprotein Receptor 1	(LOX1)	12	rs11053646	G/C	G	6	8	517	579
MyD88-adapter-like	(MAL)	11	rs8177374	T/C	T	15	17	507	550
Multi drug resistance	(MDR1) [e]	7	rs1045642	C/T	C	47	46	460	485
Nucleotide-binding oligomerization domain containing 2	(NOD2)	16	rs2066844	C/T	T	5	5	592	595
			rs2066845	C/G	G	2	1	592	595
			rs2066847	-/ C	C	3	3	592	595
Tumour necrosis factor	(TNF)	6	rs1800629	G/A	A	15	15	592	587
Tumour necrosis factor receptor 2	(TNFRSF1B)	1	rs1061622	T/G	G	24	23	618	637
Vitamin D receptor	(VDR)	12	rs731236	T/C	C	36	40	630	659
			rs7975232	C/A	C [f]	49	47	649	676

[a] Chromosome number. [b] Minor allele. [c] Minor allele frequency. [d] Available sample size for patient's or donor's SNP.
[e] Excluded from the statistical analysis because of high percentage of missing genotypes.
[f] Minor allele differs between patients (allele A) and donors (allele C).
[g] Other non-HLA polymorphisms were from genes: IFNG [h], MDR1(three allelic)[e], GCR(haplotype)[e] and IL10(haplotype)[i]
[h] Microsatellite in gene IFNG, located in chromosome 12. Alleles in patients were: 1 (<1%), 2 (47%), 3 (46%), 4 (4%), 5 (2%), 6 (<1%); the frequencies were similar in donors.
[i] Haplotype IL10 from SNPs above: rs1800896, rs1800871 (not genotyped due to complete linkage disequilibrium with the other two SNPs), and rs1800872. Haplotypes in patients/donors were: ACC (30%/28%), ATA (23%/25%), GCC (46%/46%), G_A (<1%/<1%).

Table 3. Cox regression estimates for the top ten polymorphisms associated with overall survival, in addition to the EBMT-score, ordered according to p-value for the LRT [c].

Genes from Patient (P-), Donor (D-)	Polymorphism [a]	Hazard ratio [b]	Confidence Interval (95%) [b]	LRT P value [c]	R^2 [d]	MAF % [e]
D-IL10	genotype(ACC/ACC)	0.48	0.29 - 0.80	**0.002**	0.025	28
P-MAL	rs8177374(T) add.	1.33	1.05 - 1.68	**0.021**	0.010	15
D-IL6	rs1800795(C) add.	1.19	1.03 - 1.38	**0.021**	0.007	39
P-ESR1	rs2234693(C) dom.	1.31	1.03 - 1.67	**0.027**	0.014	46
D-IL13	rs1800925(T) rec.	0.46	0.20 - 1.03	**0.032**	0.033	18
P-GCR	rs33388(T) add.	0.83	0.70 - 0.99	**0.035**	0.003	44
D-IL4	rs2243250(T) rec.	0.49	0.22 - 1.10	0.052	0.000	16
D-IL6	rs1800797(A) add.	1.18	1.00 - 1.39	0.056	0.001	39
P-IL10	rs1800872(A) rec.	0.63	0.38 - 1.05	0.057	0.010	24
D-IL10	rs1800896(G) add.	1.16	0.99 - 1.35	0.063	0.000	46

[a] Polymorphisms were ordered from lowest to highest p value for likelihood ratio test (LRT). SNP rs numbers are followed by the minor allele and the respective genetic model: additive (add.), dominant (dom.), or recessive (rec.).

[b] Hazard ratio for the polymorphism and associated 95% confidence intervals were based on the multiple Cox model including the EBMT-score and the polymorphism.

[c] P value for the likelihood ratio test comparing a multiple model including the EBMT-score and the polymorphism with respect to a model including only the EBMT-score. In bold are p-value≤0.05.

[d] R^2: Estimated gain in prediction of model containing EBMT-score and the polymorphism vs model containing EBMT-score alone.

[e] Minor allele or haplotype frequency of the respective polymorphism tested for association with overall survival.

Table 4. Cox regression model with multiple polymorphisms and EBMT-score for overall survival (n=419)

Factors [a]		Hazard ratio	Confidence interval (95%)	p-value	Reg. Coef. [b]	Scaled Coef. [c]	Risk score value [d]
EBMT-score		1.22	1.12 - 1.33	<0.001	0.20	1	EBMT-score: 0-7 [e]
D-IL10	genotype(ACC/ACC)	0.49[f]	0.26 - 0.89	0.020	-0.72	4	ACC/ACC:0, others:4
P-MAL	rs8177374(T) add.	1.34[f]	1.04 - 1.74	0.026	0.29	1	CC:0, CT:1, TT:2
P-ESR1	rs9340799(G) dom.	1.52[f]	1.15 - 2.01	0.003	0.42	2	AA:0, (AG or GG):2
D-IL6	rs1800795(C) add.	1.29[f]	1.07 - 1.55	0.007	0.25	1	GG:0, GC:1, CC:2

[a] Patients (P-) or Donors (D-). For the SNPs listed, the rs number is followed by the minor allele and the respective genetic model: additive (add.), dominant (dom.), or recessive (rec.).

[b] Regression coefficients from the Cox model, equivalent to the log of the hazard ratio.

[c] Scaled coefficient was obtained by dividing the regression coefficients by that of the EBMT-score and rounding to the nearest integer.

[d] Risk score values for polymorphisms contributing to the EBMT-genetic-score was derived using the scaled coefficients. (Supplementary Section B)

[e] Range of score values for EBMT-score[3].

[f] Note the effect size, either protective or at risk, is larger than that of EBMT-score.

Table 5. Cox regression model of overall survival with the EBMT-genetic-score using five categories (n=419) and its correlation with the EBMT-score.

Cox regression model				Correlation as number of patients with both risk scores							
EBMT-genetic-score [a]	Hazard ratio	Confidence Interval (95%)	p-value	EBMT-genetic-score	EBMT-score [0-1]	2	3	4	5	6	7	Total
1 –6[b]	1.00	- -	-	1 –6[a]	19	13	2	4	2	0	0	40
7 – 9	2.46	1.23 - 4.93	0.011	7 – 9	29	40	46	24	9	2	0	150
10	3.62	1.77 - 7.42	<0.001	10	1	11	23	29	9	5	0	78
11 – 12	4.67	2.33 - 9.33	<0.001	11 – 12	0	3	14	45	34	14	4	114
13 – 15	7.57	3.58 - 16.00	<0.001	13 – 15	0	0	0	4	10	21	2	37

[a] EBMT-genetic-score is the sum of the score values of the EBMT-score and of the score values of the four polymorphisms (Supplementary Section B).

[b] Reference group.

Table 6. Prediction of overall survival based on the EBMT-score and the EBMT-genetic-score, (n=419)

Cohort	n	C-Index [a]		Concordance	R^2 [c]
		EBMT-score	EBMT-genetic-score	EBMT-genetic-score vs EBMT-score P-value [b]	EBMT-genetic-score vs EBMT-score
Whole cohort	419	0.590	0.630	<0.001	10.8%[d]
Up to 2000	173	0.597	0.646	0.0008	1.3%[e]
After 2000	246	0.586	0.616	0.0066	25.2%[e]

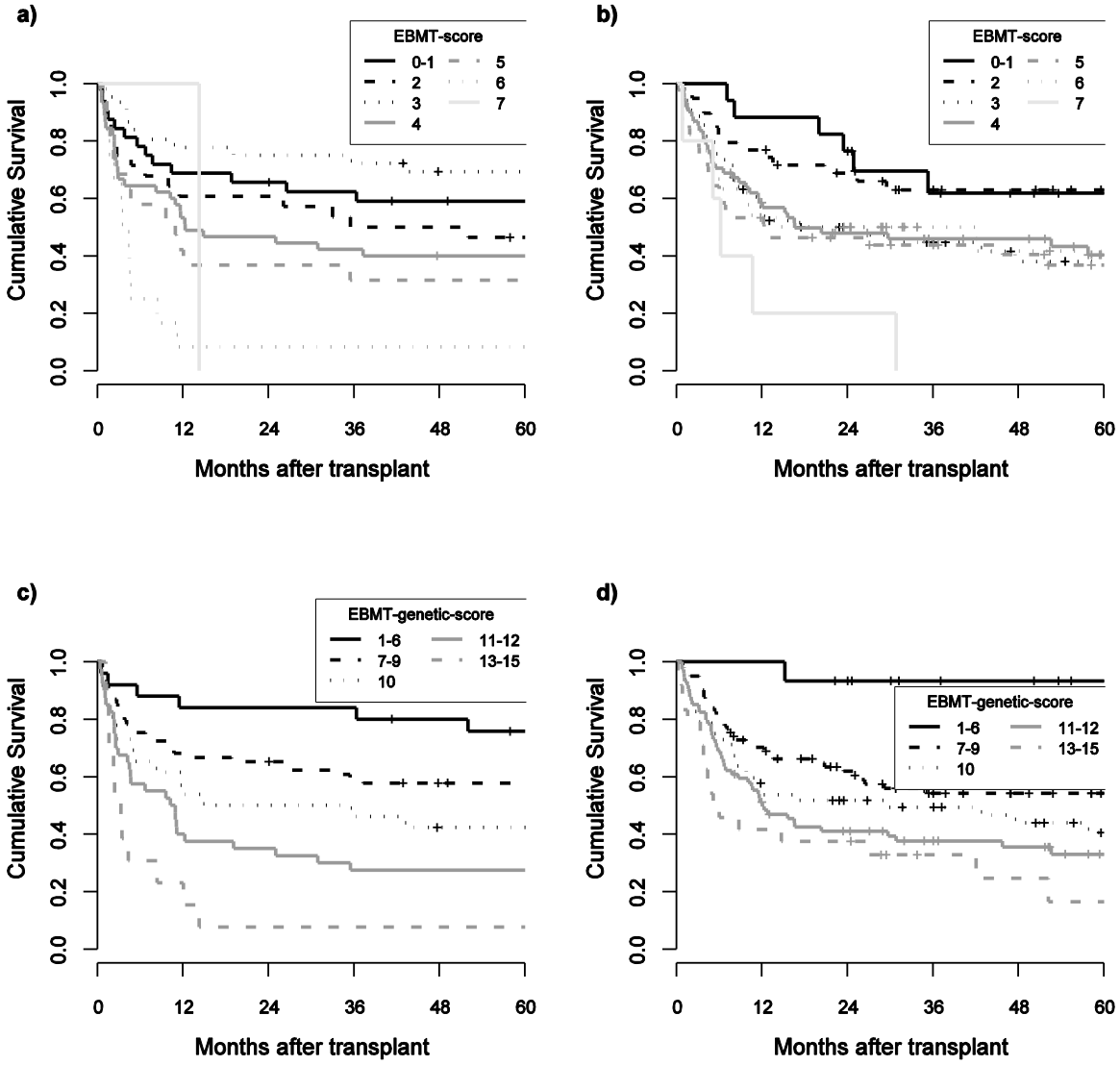
[a] Concordance Index

[b] P-value of U-Statistic[40] testing for the fraction of pairs for which EBMT-genetic-score is more in agreement with survival than the EBMT-score.

[c] Gain in prediction for overall survival[41].

[d] Gain in prediction for overall survival utilising 0.632 estimator that involves a bootstrap procedure.

[e] Gain in prediction for overall survival when whole cohort is used as a training set – utilising sub-cohort up to or after 2000 as a validation set.


a)
EBMT-score
0-1
2
3
4
5
6
7
Cumulative Survival
0.0
0.2
0.4
0.6
0.8
1.0
0
12
24
36
48
60
Months after transplant
b)
EBMT-score
0-1
2
3
4
5
6
7
Cumulative Survival
0.0
0.2
0.4
0.6
0.8
1.0
0
12
24
36
48
60
Months after transplant
c)
EBMT-genetic-score
1-6
7-9
10
11-12
13-15
Cumulative Survival
0.0
0.2
0.4
0.6
0.8
1.0
0
12
24
36
48
60
Months after transplant
d)
EBMT-genetic-score
1-6
7-9
10
11-12
13-15
Cumulative Survival
0.0
0.2
0.4
0.6
0.8
1.0
0
12
24
36
48
60
Months after transplant

Supplementary Table 1.

Top ten polymorphisms from the whole cohort that remained as top polymorphisms in the subgroup analyses for association with overall survival.

Genes from Patient (P-), Donor (D-)	Polymorphism [b]	P value level[a] in subgroups											
		MUD	SIB	T-depl	No T-depl	RIC	MC	AL	CML	Lymph	Other Dx	≤2000	>2000
	Sample size	358	385	256	487	259	484	377	169	110	87	296	447
D-IL10 [c]	genotype(ACC/ACC)	3	1	3	1	1	1	2	2	1	1	1	1
P-MAL	rs8177374(T)	-	1	-	1	1	-	1	-	-	-	1	-
D-IL6	rs1800795(C)	1	-	-	1	-	1	-	1	-	-	1	-
P-ESR1	rs2234693(C)	-	1	-	1	-	1	-	2	1	-	2	-
D-IL13	rs1800925(T)	3	-	-	1	-	1	1	-	-	-	-	2
P-GCR	rs33388(T)	1	-	-	-	-	1	1	3	-	-	2	3
D-IL4	rs2243250(T)	-	2	-	2	-	1	2	-	2	2	-	-
D-IL6	rs1800797(A)	1	-	-	2	-	2	-	1	-	1	1	-
P-IL10	rs1800872(A)	-	3	-	1	3	1	-	-	-	2	-	-
D-IL10	rs1800896(G)	3	-	-	2	2	2	2	-	1	-	-	1

MUD: matched unrelated donors, SIB: sibling donors, T-depl: T-cell depletion, no T-depl: no T-cell depletion, RIC: reduced intensity conditioning, MC: myeloablative conditioning, AL: acute leukaemia, CML: chronic myeloid leukaemia, Lymph: lymphoma, Other Dx: other diagnoses (different than AL, CML, or Lymph), ≤2000: transplant up to and including 2000, >2000: transplant after 2000.

[a] P value from the likelihood ratio test: 1. <0.05, 2. [0.05, 0.10], 3. >0.10, - polymorphism was not among the top ten.

[b] Top polymorphisms from the whole cohort (Table 3). SNP rs numbers are followed by the minor allele.

[c] Haplotype D-IL10 was within the top ten polymorphisms with haplo-genotype (ACC/ACC). This polymorphism appears within the top ten list with haplotype GCC in subgroups MUD, AL, and >2000; and with haplotype ATA in subgroups CML and Other Dx.

Supplementary Table 2.

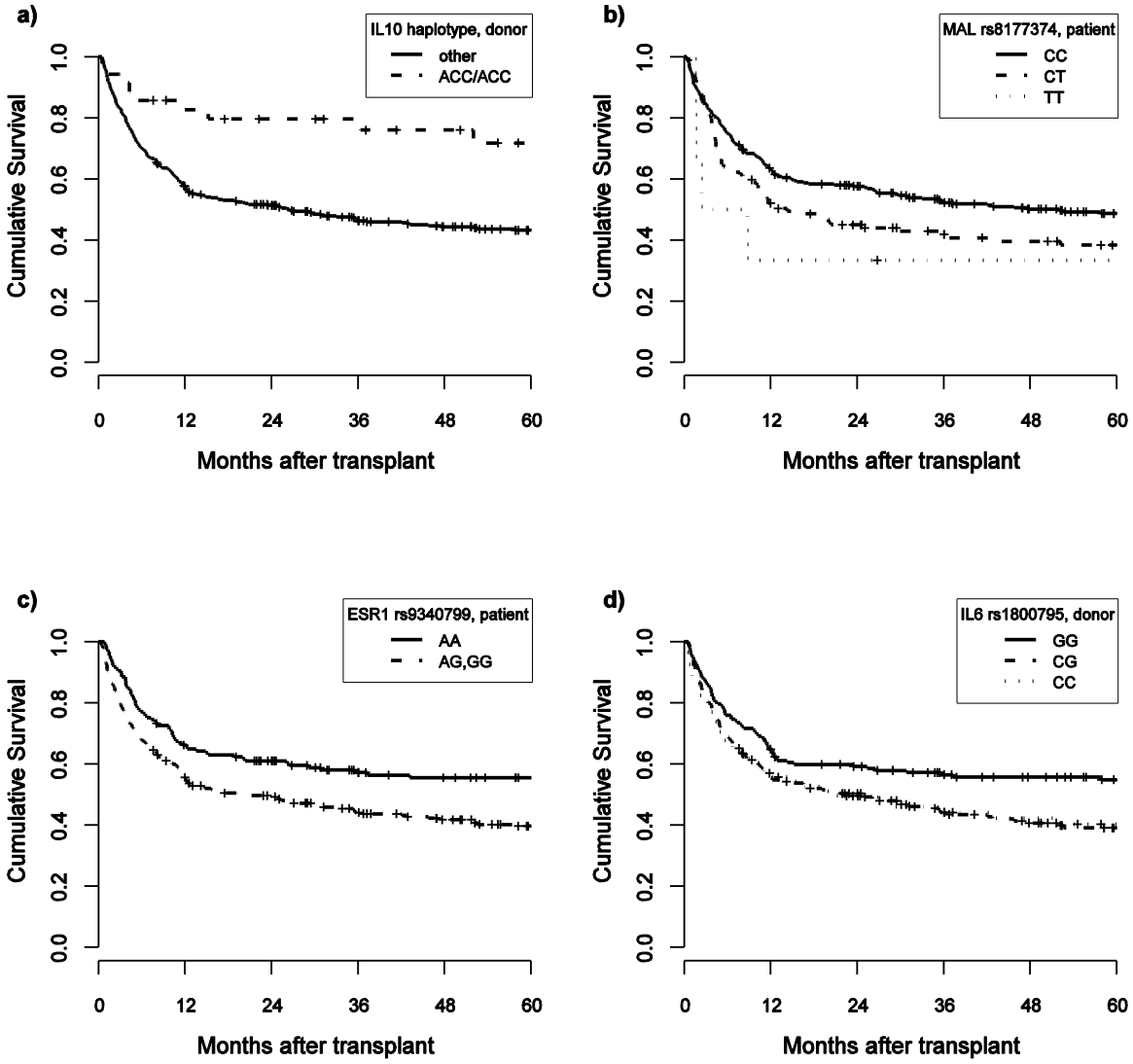
P-values for the association of selected polymorphisms with different causes of death (n=419), using a Cox regression model including a single polymorphism and the EBMT-score.

Genes from Patient (P-), Donor (D-)	Polymorphism [a]	P-values from LRT [b]			
		Overall death	Cause of death		
			Relapse	Infection	GvHD
D-IL10	genotype(ACC/ACC)	0.013	0.003	0.503	0.494
P-MAL	rs8177374(T) add.	0.023	0.019	0.568	0.026
P-ESR1	rs9340799(G) dom.	0.006	0.050	0.086	0.118
D-IL6	rs1800795(C) add.	0.006	0.100	0.003	0.680

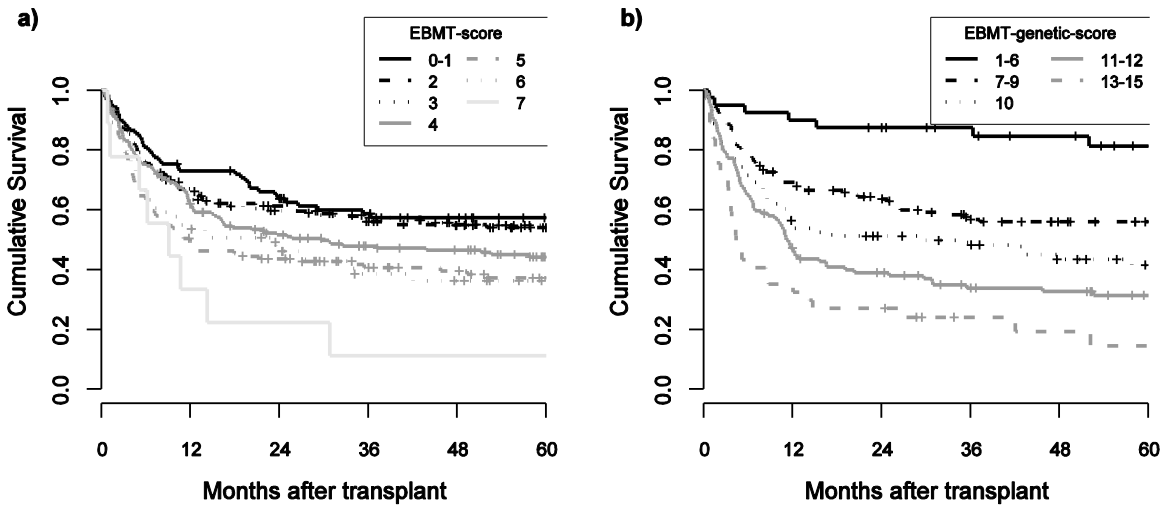
[a] Patients (P-) or Donors (D-). For the SNPs listed, the rs number is followed by the minor allele and the respective genetic model: additive (add.) or dominant (dom.).

[b] P-value for the likelihood ratio test comparing a Cox model including the EBMT-score and the polymorphism with respect to a model including only the EBMT-score.

Supplementary Fig. 1:
KM survival curves of HSCT patients for polymorphisms in the Cox model (n=419)

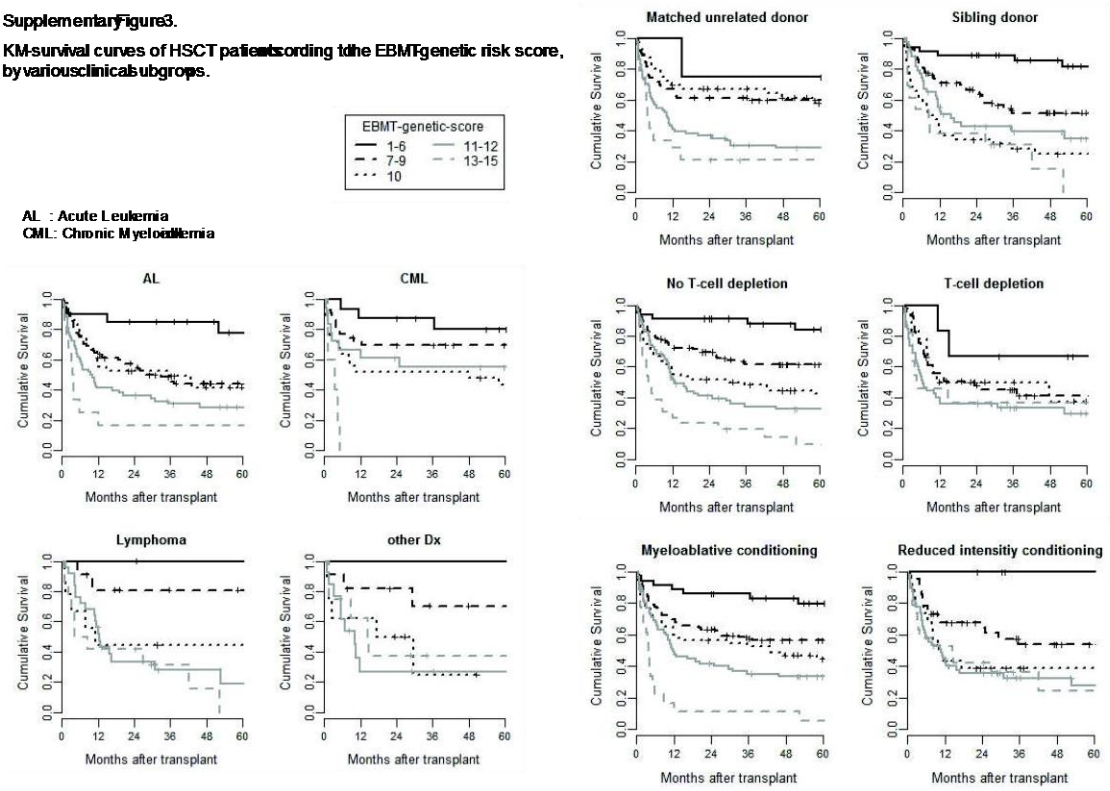


Supplementary Fig.2:
KM survival curves of HSCT patients for clinical and genetic risk scores, a) n=743, b) n=419.



Supplementary Figure 3.

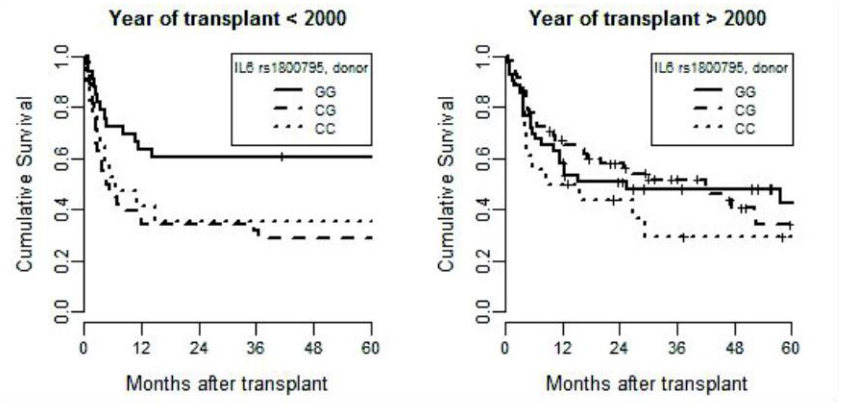
KM-survival curves of HSCT patients according to the EBMT-genetic risk score, by various clinical subgroups.



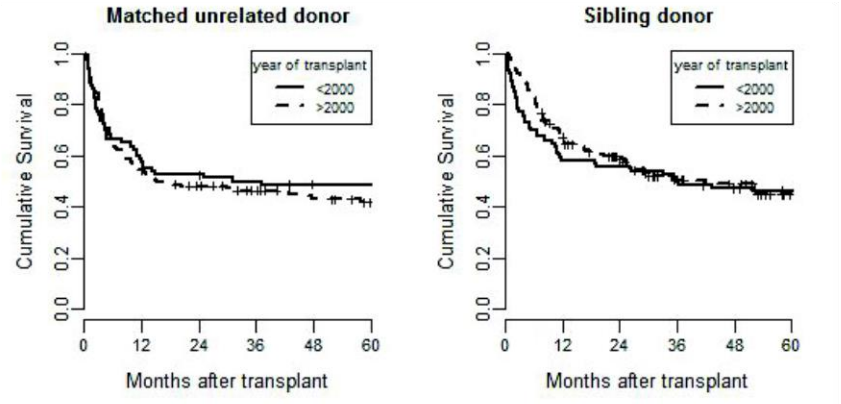
Supplementary Figure 4.

KM-survival curves comparing overall survival of
a) patients whose donors carry IL6 rs1800795, by year of transplant
b) patients receiving HSCT before and after year 2000, by type of donor

a)



b)



Supplementary Section A

Biallelic SNPs were evaluated under the additive, dominant and recessive modes of inheritance. As an example of coding, consider 3 genotypes AA, AT and TT and T as the minor allele: dominant coding is (AT or TT)=1, AA=0; recessive coding is: TT=1, (AT or AA)=0; additive coding is: AA=0, AT=1, TT=2. For the IL-10 haplotype analysis, there are three possible haplotypes GCC, ATA and ACC – the haplo-genotypes of IL10 could therefore be GCC/GCC; GCC/ATA; GCC/ACC; ATA/ATA; ATA/ACC or ACC/ACC – thus binary coding for comparing a particular haplo-genotype to others was, for example, ACC/ACC=1, others=0.

Supplementary Section B

The EBMT-genetic-score is derived after summing the risk score elements of the EBMT-score and genetic factors (final column, Table 4). For example, a patient with EBMT-score=3, presence of IL-10 (ACC/ACC) in the donor (contribution 0), presence of TT for MAL (patient) (contribution 2), presence of AG or GG for ESR1 (patient) (contribution 2) and presence of CC for IL-6 (donor) (contribution 2) would have EBMT-genetic-score= 3+0+2+2+2= 9.

Supplementary Section C

$$R^2 = \left(1 - \frac{1 - R^2_{.632,EBMT-genetic-score}}{1 - R^2_{.632,EBMT-score}}\right) \quad \text{x } 100$$

where

$$R^2_{.632,EBMT-genetic-score} = 1 - \frac{\bar{\bar{Err}}_{.632,EBMT-genetic-score}}{\bar{\bar{Err}}_{.632,null}}$$

and

$$R^2_{.632,EBMT-score} = 1 - \frac{\widetilde{Err}_{.632,EBMT-score}}{\widetilde{Err}_{.632,null}}$$

$\widetilde{Err}_{.632}$ incorporates a bootstrap cross-validation estimator. In this case, the bootstrapping cross-validation procedure is implemented by sampling a data set (of size n) with replacement, this data set is a 'bootstrap sample' and acts as a training sample to calculate a model. The model is then tested on the data not used in the training sample – i.e. these data act as a validation set. The observed and predicted values are compared. This procedure is repeated many times and an average found – this is the 'bootstrap cross-validation estimate of the prediction errors'.

For $\widetilde{Err}_{.632,EBMT-genetic-score}$ and $\widetilde{Err}_{.632,EBMT-score}$, the predicted values are calculated using the Cox model with the indicated factors included; for $\widetilde{Err}_{.632,null}$ the predicted values are calculated using a model without factors,(corresponding to the baseline survival risk).

$\widetilde{Err}_{.632}$ is small when there is good agreement between observed and predicted values and $\widetilde{Err}_{.632}$ is large when there is poor agreement between observed and predicted values. Therefore a value of 0% for R^2 means that the model with EBMT-score and genetic factors and the model with EBMT-score alone have equivalent predictive ability; a value of 100% for R^2 means that the model with EBMT-score and genetic factors has perfect predictive ability (i.e. the predicted and observed outcome always agree).